<table>
<thead>
<tr>
<th>Title</th>
<th>Functional Evolution in Orthologous Cell-encoded RNA-dependent RNA Polymerases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Qian, Xinlei; Hamid, Fursham M.; El Sahili, Abbas; Darwis, Dina Amallia; Wong, Yee Hwa; Bhushan, Shashi; Makeyev, Eugene V.; Lescar, Julien</td>
</tr>
<tr>
<td>Date</td>
<td>2016</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/40997">http://hdl.handle.net/10220/40997</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2016 The American Society for Biochemistry and Molecular Biology. This is the author created version of a work that has been peer reviewed and accepted for publication by Journal of Biological Chemistry, The American Society for Biochemistry and Molecular Biology. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<a href="http://dx.doi.org/10.1074/jbc.M115.685933">http://dx.doi.org/10.1074/jbc.M115.685933</a>].</td>
</tr>
</tbody>
</table>
Manuscript Title: Functional Evolution in Orthologous Cell-encoded RNA-dependent RNA Polymerases

Manuscript No: JBC/2015/685933 [R1]

Manuscript Type: Regular Paper

Date Submitted by the Author: 4 Feb 2016

Complete List of Authors: Xinlei Qian, Fursham M. Hamid, Abbas El Sahili, Dina Amallia Darwis, Yee Hwa Wong, Shashi Bhushan, Eugene V. Makeyev, and Julien Lescar

Keywords: electron microscopy (EM); protein evolution; RNA interference (RNAi); RNA polymerase; X-ray crystallography; RNA-dependent RNA polymerase
Functional evolution in orthologous cell-encoded RNA-dependent RNA polymerases

Xinlei Qian1, Fursham M. Hamid1, Abbas El Sahili1, Dina Amallia Darwis1, Yee Hwa Wong1, Shashi Bhushan1, Eugene V. Makeyev1,2*, and Julien Lescar1,3*

1 Division of Structural Biology and Biochemistry, School of Biological Sciences, Nanyang Technological University, 138673, Singapore
2 MRC Centre for Developmental Neurobiology, King’s College, London, SE1 1UL, United Kingdom
3 UPMC UMRS CR7 - CNRS ERL 8255-INSERM U1135 Centre d’ Immunologie et des Maladies Infectieuses. Centre Hospitalier Universitaire Pitié-Salpêtrière, Faculté de Médecine Pierre et Marie Curie, Paris, France

Running title: Functional evolution of fungal RdRPs

* To whom correspondence should be addressed. Tel: +65 65869706; Fax: +65 67913856; Email: julien@ntu.edu.sg; E-mail: eugene.makeyev@kcl.ac.uk

Present Address: Xinlei Qian, Singapore-MIT Alliance for Research and Technology, 138602, Singapore

Keywords: RNA-dependent RNA polymerase; small RNAs; QDE-1 orthologs; X-ray crystallography, electron microscopy; evolution.

ABSTRACT

Many eukaryotic organisms encode more than one RNA-dependent RNA polymerase (RdRP) that likely emerged as a result of gene duplication. Such RdRP paralogs often participate in distinct RNA silencing pathways and show characteristic repertoires of enzymatic activities in vitro. However, to what extent members of individual paralogous groups can undergo functional changes during speciation remains an open question. We show that orthologs of QDE-1, an RdRP component of the quelling pathway in Neurospora crassa, have rapidly diverged in evolution at the amino acid sequence level. Analyses of purified QDE-1 polymerases from N. crassa (QDE-1Ncr) and related fungi, Thielavia terrestris (QDE-1Tie) and Myceliophthora thermophila (QDE-1Mth), show that all three enzymes can synthesize RNA but the precise modes of their action differ considerably. Unlike their QDE-1Ncr
counterpart favoring processive RNA synthesis, QDE-1^Te and QDE-1^Ncr produce predominantly short RNA copies via primer-independent initiation. Surprisingly, a 3.19 Å-resolution crystal structure of QDE-1^Te reveals a quasi-symmetric dimer similar to QDE-1^Ncr. Further electron microscopy analyses confirm that QDE-1^Te occurs as a dimer in solution and retains this status upon interaction with a template. We conclude that divergence of orthologous RdRPs can result in functional innovation while retaining overall protein fold and quaternary structure.

INTRODUCTION

Eukaryotic cells widely use small RNA (sRNA) guides to limit proliferation of viruses and transposable elements, maintain proper chromosomal structure and control endogenous gene expression in a sequence-specific manner (1-10). sRNA pathways typically require RNase III-like and PIWI/PAZ proteins that process completely or partially double-stranded (ds) RNA precursors and recruit sRNA products of this reaction into functional silencing complexes (11-13). In fungi, plants, protozoans and some metazoans, sRNA production often depends on template-dependent RNA synthesis catalyzed by cell-encoded RNA-dependent RNA polymerases (RdRPs) (14,15).

Many RdRPs contribute to maintenance and amplification of gene silencing signals initiated by primary sRNAs originating from viral genomes, endogenous sequences or dsRNAs experimentally delivered into a cell (16-23). In such cases, long single-stranded (ss) RNA targets of an initial round of silencing become templates for RdRP-catalyzed RNA synthesis ultimately giving rise to secondary sRNAs. Moreover, some RdRPs may trigger a silencing response with no apparent need for pre-existing sRNAs (22-27). This may involve RdRP recruitment to transcripts with aberrant or unusual molecular features generated by other RNA polymerases (23,24,28,29).

In the filamentous fungus Neurospora crassa, silencing, or "quelling", of transgenic arrays relies on RdRP QDE-1, PIWI/PAZ protein QDE-2, DNA helicase QDE-3 and several other components including the ssDNA-binding protein RPA (2). Interestingly, QDE-1 may trigger quelling by producing long aberrant ssRNA copies (aRNAs) of ssDNA intermediates that frequently arise in tandem-duplicated genomic sequences (30-32). This reaction depends on QDE-3 and RPA and is stimulated by DNA damage. QDE-1 can subsequently use its aRNA product as a template to produce dsRNA intermediates. These are subsequently converted into sRNAs that associate with QDE-2 and target complementary sequences. Thus, at least some RdRPs may initiate gene silencing de novo.

In line with their diverse biological functions, individual RdRPs and their protein complexes isolated from various sources show different enzymatic properties in vitro. These include polymerase-specific preferences between ssRNA vs. ssDNA templates and primer-dependent vs. primer-independent initiation of RNA synthesis (33-37). Some RdRPs can also function as template-independent terminal transferases (34,35). Notably, two distinct template-dependent polymerization modes have been described for previously studied RdRPs: (a) processive synthesis of long ds products, which is typically initiated at or close to the 3' end of an ss template using either primer-independent or so-called "back-priming" mechanisms and (b) non-processive synthesis of sRNA copies initiated in a primer-independent manner at internal positions of an ss template (33).

Different RdRPs appear to utilize the two modes with markedly different efficiencies. For example, Rdr1 from S. pombe or RDR6 from Arabidopsis efficiently synthesize long products (34,38), whereas
RRF-1 from *C. elegans* specializes in production of sRNAs (39). These biochemical differences are consistent with *in vivo* evidence: long dsRNAs generated by Rdr1, RDR2, and RDR6 must be processed by Dicer/RNAseIII-like endoribonucleases to generate functional small interfering (si) RNAs while sRNA products of RRF-1 apparently do not require further processing for their secondary siRNA function (16,19,20).

Purified QDE-1 from *N. crassa* can use both primer-independent and “back-priming” mechanisms *in vitro* (33,35). Crystal structure of QDE-1 catalytic fragment – thus far the only known structure of a cell-encoded RdRP – suggests that this enzyme is a homodimer with the two subunits adopting either “closed” or “open” conformation (40). It has been proposed that the two structurally distinct conformations may help this remarkably versatile enzyme choose between different activities (40). However, in the absence of structural information for corresponding enzyme-substrate complexes, whether QDE-1 in fact remains a dimer upon template binding is unknown. Moreover, it remains to be seen if other RdRPs can form homodimers, an important question given that at least some RdRPs behave as monomers in solution (36).

On a more fundamental level, how new functional properties evolve in the RdRPs – and in other protein families for that matter – is poorly understood. Gene duplication followed by paralog divergence is a major driving force in protein evolution (41,42) and it clearly contributed to RdRP diversification. Indeed, many species encode more than one distinct RdRP, with three paralogous genes present in *N. crassa* (QDE-1, SAD-1 and RRP-3), four in *C. elegans* and six in *Arabidopsis* (14). The last eukaryotic common ancestor might have contained three functionally distinct RdRPs giving rise to the α, β and γ branches of the RdRP genealogy, an arrangement that was further modified by lineage-specific gene duplications and losses (43).

Species-specific members of individual paralogous groups, referred to as orthologs, are typically assumed to have similar biological activities (44). However, it has been alternatively proposed that divergence of orthologous sequences might frequently result in acquisition of novel functional properties (45). Until recently, it has been difficult to investigate these possibilities experimentally since genomes of just a few distantly related model organisms have been sequenced completely. Here we took advantage of the increasing number of whole-genome sequences available for fungal species and examined evolutionary trends in QDE-1 orthologs using phylogenetic, biochemical and structural approaches.

**EXPERIMENTAL PROCEDURES**

**Phylogenetic analyses**—Amino acid sequences of fungal polymerases were downloaded from OrthoDB (46) and aligned using MUSCLE (47). Phylogenetic trees were constructed in MEGA6 (48) by computing evolutionary distances using Poisson correction and inferring evolutionary history by the Neighbor-Joining method. Tree topology was tested using bootstrapping. Amino acid sequence conservation profiles were plotted in EMBOSS/plotcon using a 50-aa sliding window (49). Protein structures were color-coded according to interspecies conservation using Chimera (50). Chimera was also used to predict ligand positions in *N. crassa* QDE-1 apoenzyme [PDB 2J7N; (40)] based on the known structure of the pol-II elongation complex [PDB 1R9T; (51)].

**Protein expression and purification**—Recombinant proteins were expressed and purified as described elsewhere (52,53). Briefly, synthetic open reading frames (ORFs) encoding catalytic fragments of QDE-1*Te* and QDE-1*Mb* were obtained from Genscript and
the sequence encoding catalytic fragment of QDE-1
(src (“QDE-1ΔN”) was amplified from pEM55 (33). Catalytically inactive QDE-1
Mth D607A mutant with the DYDGD motif substituted with AYDGD was prepared by QuikChange mutagenesis (Agilent). The ORFs were amplified using primers shown in Table 1, cloned into the pFB-LIC-Bse (a gift from Opher Gileadi; Addgene plasmid # 26108) using ligation independent cloning (54), and subsequently transformed into DH10Bac (Life Technologies) to produce the recombinant bacmids. Viral stocks generated by introducing the bacmids into Sf9 insect cells were further amplified and used to infect Sf9 cells for large-scale protein expression. Virus-infected cells were harvested by centrifugation at 4,000×g for 15 min at 4°C. Cell pellets were resuspended in 20 mM HEPES-NaOH, pH 8.0, 5 mM imidazole, 300 mM NaCl, 5% (v/v) glycerol with complete EDTA-free protease inhibitors (Roche) and subjected to sonication. Soluble fractions were isolated by centrifugation at 50,000×g for 30 min at 4°C and incubated with His-tag purification resin (Roche) for 1 hour at room temperature. Non-specifically bound proteins were eluted by 20 mM HEPES-NaOH, pH 8.0, 15 mM imidazole, 300 mM NaCl, 5% (v/v) glycerol and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP). His tag-containing proteins were eluted with 20 mM HEPES-NaOH, pH 8.0, 500 mM imidazole, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP. Fractions containing the protein were concentrated using a 100 kDa cutoff concentrator (Sartorius) and further purified by size exclusion chromatography using Superdex 200 (GE Healthcare) pre-equilibrated with 20 mM HEPES-NaOH, pH 8.0 and 0.5 mM TCEP additionally containing 150 mM NaCl and 5% (v/v) glycerol (QDE-1
Nε), 300 mM NaCl and 5% (v/v) glycerol (QDE-1
Nε) or 500 mM NaCl and 10% (v/v) glycerol (QDE-1
Mth and QDE-1
Mth D607A). Eluted proteins were concentrated and stored at -80°C. Chemicals were from Sigma-Aldrich unless stated otherwise.

Protein thermostability assay—Protein thermostability was determined by monitoring temperature-induced fluorescence changes, as described elsewhere (55). Purified proteins were incubated at 1 mg/mL in the gel filtration buffer with 1000-fold diluted SYPRO Orange stock (Life Technologies) in 96-well PCR plates (Bio-Rad) sealed with Optical Sealing Tape (Bio-Rad). Fluorescence was measured using an iCycler iQ5 Real-Time PCR Detection System (Bio-Rad) with excitation and emission wavelengths set at 490 and 575 nm, respectively. Temperature was increased from 20 to 90°C with 1°C increments and the mixture was incubated for 12 seconds at each temperature value. Protein melting temperatures were calculated using the iQ5 Optical System Software, Version 2.1 (Bio-Rad).

RNA polymerase assays—A synthetic single-stranded DNA (ssDNA) template (5’-CTGACTGCTTCTGTTTCTTTCTCTCCCCTTTTTTCTCATGCCCCACACCACACGTTCTTCTTTGCTGCTCTACCCTGGACAATTAAATCATCGGCA-3’) was synthesized by Sigma. A single-stranded RNA (ssRNA) template corresponding to the s+ transcript of bacteriophage φ6 was produced from pLM659 (56) linearized with SmaI using the mMessage mMachine T7 transcription kit (Ambion). QDE-1 assays were carried out in 50 mM HEPES-NaOH, pH 7.8, 0.1 mM EDTA, 2% (v/v) Triton X-100, 100 mM NH4OAc, 2 mM MgCl2, 0.1 units/µL of recombinant RNasin (Promega), 0.4 mM each of ATP, CTP and GTP, 0.2 mM UTP and 25 µCi/µL of [α-32P]UTP (PerkinElmer Life Sciences). Reactions were initiated by adding a corresponding QDE-1 polymerase to the final concentration of 0.25 µg/µL followed by 1-hour incubations at 30-65°C. Reaction products were separated as described (33) using native or denaturing formaldehyde-containing agarose gel electrophoresis or urea-
Containing polyacrylamide gel electrophoresis (PAGE) and analyzed by phosphorimaging (Typhoon Trio, GE Healthcare). RNase protection assay was carried out by stopping the polymerase reactions with 250 mM NH₄OAc and 10 mM EDTA followed by incubation with various concentrations of RNase ONE (Promega) for 1 hr at 30°C. The reaction products were analyzed by formaldehyde-containing agarose gel electrophoresis as described above. To quantify radioactivity incorporated into newly synthesized RNA products, reaction mixtures were spotted onto Whatman 3MM paper pre-treated with 10% trichloroacetic acid (TCA) solution, washed in ice-cold 10% TCA solution for 10 min, rinsed twice with ice-cold 10% TCA solution and once with 95% ethanol and allowed to air-dry. To determine the sum of incorporated and non-incorporated radioactivity, samples were spotted on Whatman 3MM paper and air-dried with no additional treatment. Radioactive signals were then analyzed by phosphorimaging.

**Filter binding assay**—Purified QDE-1ᴺcr, QDE-1¹ᵗᵉ or bovine serum albumin (BSA) control were incubated at 1 μM in the presence of binding buffer (50 mM HEPES-NaOH, pH 7.8, 0.1 mM EDTA, 2% (v/v) Triton X-100, 100 mM NH₄OAc and 2 mM MgCl₂) and 0.03 to 10 μM of 5'-CTTACTTGATGGACATTT-3' ssDNA oligonucleotide labeled at the 5' end using [γ-³²P]-ATP (Perkin Elmer) and T4 polynucleotide kinase (NEB). Following a 10-min incubation at 20°C, mixtures were passed through nitrocellulose membrane prewashed with the binding buffer using a Bio-Rad slot blot apparatus. The labeled RdRP-ssDNA complex retained on the membrane was quantified using phosphorimaging (Typhoon Trio, GE Healthcare).

**Protein crystallization, diffraction data collection and structure determination**—Crystals of the QDE-1¹ᵗᵉ catalytic fragment were obtained at 30°C by mixing 2 μL of the protein solution at 2 mg/mL with 1 μL crystallization solution (100 mM Tris-HCl at pH 8.0, 75 mM NaCl, 10% (w/v) PEG 10,000, 5 mM MgCl₂ and 6 mM spermine). Crystals were transferred to a drop containing the crystallization solution supplemented with 10% (v/v) glycerol and incubated at 12 °C overnight. Before freezing, crystals were dehydrated at RT in two steps of 15 min each in the crystallization solution supplemented with 20% and 30% glycerol. X-ray diffraction data were collected to 3.19 Å resolution at 100 K at the PXIII beamline of the Swiss Light Source, Villigen, Switzerland using a Pilatus 6M detector (Dectris). The crystal belonged to the P2₁ space group with the following unit cell dimensions: a = 84.23 Å, b = 165.84 Å, c = 173.83 Å and β=90.10°. The data collection and structure refinement parameters are listed in Table 2. The structure was determined by molecular replacement using the known structure of the QDE-1ᴺcr catalytic fragment (Protein Data Bank code 2J7N; (40)) as a search probe. The model was built interactively using Coot (57) and the structure was refined using REFMAC from the CCP4 package, with tight ncs restraints between the four independent monomers (each monomer was considered as a group) with individual atom isotropic temperature factors and TLS refinement (58). Each monomer of the QDE-1¹ᵗᵉ fragment used for crystallization contains 1034 residues, including the His tag and the TEV protease cleavage site. Of these, 924/922 could be built in dimer A/B and 922/922 in dimer C/D in the two non-crystallographic dimers of the asymmetric unit (Table 2). Missing residues in the model belong to the N- and C- termini and flexible loops (Table 2). Since QDE-1¹ᵗᵉ dimer A/B is better ordered than dimer C/D in the electron density map, we use it for subsequent comparisons.

**Electron microscopy (EM)**—Purified QDE-1¹ᵗᵉ was diluted to 10 μg/mL with 100 mM Tris-HCl, pH 7.5, 75 mM NaCl, 5 mM MgCl₂, and 5% (v/v) glycerol. A volume of 4 μL of
protein sample was applied to a glow-discharged carbon-coated transmission electron microscopy (TEM) grid and stained with 2% (v/v) uranyl acetate. RdRP-ssDNA substrate complex was formed by incubating 1.3 μM QDE-1<sup>Tie</sup> with 1.7 μM (i.e. ~10 higher than the apparent K<sub>d</sub> determined in the filter binding assay) DNA oligonucleotide 5’-CCTTAATTGTAG-3’ prior to transferring acutely diluted mixture to a grid and subsequently staining with uranyl acetate as above. Images were recorded at a magnification of 66,350× using a FEI T12 transmission electron microscope equipped with a 4K CCD camera (FEI) under low dose conditions. Single particles were selected and processed with the EMAN2 image-processing package (59). Initially, a total of 3,000 particles each of the QDE-1<sup>Tie</sup> apoenzyme and of QDE-1<sup>Tie</sup> pre-incubated with ssDNA were used for 2D classifications and 3D reconstructions. For both data sets, ten initial models were generated with EMAN2 and compared with the 40 Å resolution filtered crystal structure of the QDE-1<sup>Tie</sup> dimer. The initial model structurally most similar to the crystal structure was selected for further refinement for both data sets in order to obtain a 3D EM map at about 31Å resolution without imposing any symmetry constraint. To obtain a higher resolution map of QDE-1<sup>Tie</sup> pre-incubated with ssDNA, a data set of 11,992 particles was used and processed as described above which led to a “reference-free” 3D EM map at a resolution of 20 Å. In parallel, the same data set for QDE-1<sup>Tie</sup> pre-incubated with ssDNA particles was processed as described above but with the crystal structure of apo-QDE-1<sup>Tie</sup> as a reference.

RESULTS

Orthologs of Neurospora crassa QDE-1 belong to a rapidly evolving protein group—To gain insights into RdRP evolution, we analyzed corresponding genes from 40 taxonomically diverse fungi with completely sequenced genomes (Table 3 and Fig. 1). Most species encoded RdRP proteins clustering with N. crassa QDE-1, SAD-1 or RRP-3, except for three species from the Eurotiales and Hypocreales orders that had QDE-1 and SAD-1 but not RRP-3. Schizosaccharomyces pombe and Schizosaccharomyces japonicus had a single RdRP (Rdr1) related to SAD-1 and Saccharomyces cerevisiae had no RdRPs, as expected. Within each of the three groups, RdRPs clustered according to their taxonomic origin (Fig. 1). This topology suggested that the last common ancestor of fungi might have contained at least three RdRP paralogs that evolved as orthologously related lineages or were occasionally lost during speciation.

Notably, the QDE-1 orthologs tended to be connected by longer branches than their SAD-1 and RRP-3 counterparts (Fig. 1). To identify regions accounting for their apparently accelerated evolution, we examined QDE-1 conservation plot (Fig. 2). The non-catalytic N-terminal part showed extremely low conservation scores, as pointed out previously (33). However, even within the generally less divergent C-terminal part, a prominent peak of sequence conservation was detected only in the vicinity of the catalytic DYDGD motif (14,33). This contrasted with the SAD-1 and RRP-3 plots that contained substantially broader regions of relatively high conservation. Conservation was even more uniform for two pol-II subunits, RPB1 and RPB2, distantly related to cell-encoded RdRPs (40) (Fig. 2A). Quantitative analyses of amino acid substitution rates showed that QDE-1 was significantly more divergent than SAD-1, RRP-3, RPB1 and RPB2 (Fig. 2B).

Interestingly, conserved amino acid residues showed prominent clustering around the active center of the known crystal structure of the QDE-1<sup>Ncr</sup> apoenzyme interacting with one of the two catalytic Mg<sup>2+</sup> ions (Mg<sup>2+</sup>·A) [PDB 2J7N; (40)] (Fig. 2C and Fig. 3). We modeled positions of other molecules participating in RNA polymerization
including an incoming ATP monomer, the second Mg\(^{2+}\) ion (Mg\(^{2+}\)B), the template and the RNA product based on the structure of the pol-II elongation complex [PDB 1R9T; (51)]. This placed the ATP and Mg\(^{2+}\)B near the conserved surface of the nucleotide pore and the nascent 3’ end of the RNA product along with the corresponding template nucleotides near the conserved DYDGDI loop and its Mg\(^{2+}\)A ligand (Fig. 2C and Fig. 3). On the other hand, a more distal segment of the template-product duplex egressing from the active center was surrounded by substantially more divergent QDE-1 surfaces (Fig. 2C and Fig. 3). Notably, all contacts made by the template and the nascent RNA with the RPB1 and RPB2 subunits of the pol-II complex appear highly conserved in evolution (Fig. 2C). We concluded that the QDE-1 group is generally more divergent than its RdRP paralogs and DdRP relatives, and that it shows unusually strong sequence variability outside of the active center and the NTP-interacting surfaces.

QDE-1 orthologs generate markedly different combinations of long and short RNA products—To examine whether divergent QDE-1 proteins had distinct functional properties, we purified catalytic fragments of QDE-1\(^{Ncr}\) and its ortholog from *Thielavia terrestris* (QDE-1\(^{Tte}\)) from the Chaetomiaceae family distantly related to *N. crassa* (Sordariaceae family) using a standardized protocol (see Materials and Methods) and analyzed the RNA polymerase activity of these two proteins (Figs 4 and 5). QDE-1\(^{Ncr}\) is known to accept either single-stranded (ss) RNA or DNA templates and generate their continuous end-to-end RNA copies through a primer-independent (de novo) or a primer-dependent (“back-priming”) initiation mechanism (33,35). A distinct primer-independent mode allows QDE-1\(^{Ncr}\) to produce sRNA copies of internal template sequences (33). We therefore assayed QDE-1\(^{Ncr}\) and QDE-1\(^{Tte}\) RNA polymerase activities using recombinant ssRNA (2948 nt) and ssDNA (107 nt) templates and analyzed reaction products using native agarose gel electrophoresis. As expected (33), QDE-1\(^{Ncr}\) readily synthesized detectable amounts of both long and short RNA copies migrating on native gels as full-length dsRNAs and partial dsRNA species, respectively (Fig. 4A).

Similar to its *N. crassa* ortholog, the newly analyzed QDE-1\(^{Tte}\) was a highly active RNA polymerase (Fig. 4A). However, it showed a striking bias towards generating short RNA products rather than full-length copies (Fig. 4A). This activity produced readily detectable amounts of template-product duplexes migrating slower than the ssRNA template but faster than the corresponding full-length dsRNA on EtBr-stained gels (Fig. 4A). Gel autoradiography additionally revealed a prominent low-molecular-weight band likely corresponding to short RNA copies released from the template (Fig. 4A). A similarly migrating band of short RNAs was also a major reaction product in the ssDNA-programmed reactions containing QDE-1\(^{Tte}\) but not QDE-1\(^{Ncr}\) (Fig. 4B). On the other hand, both enzymes gave rise to full-length DNA/RNA template/product duplexes with comparable efficiencies (Fig. 4B).

To better understand the nature of RNA products, we analyzed ssRNA-programmed reactions by denaturing agarose gel electrophoresis (Fig. 4C). Under these conditions, short RNA copies quantitatively dissociated from the template and migrated at the expected low-molecular-weight position. This analysis additionally revealed RNA products of \(1 \times \) template length thus suggesting that – under conditions used in our RdRP assays – both QDE-1\(^{Ncr}\) and QDE-1\(^{Tte}\) can initiate end-to-end RNA synthesis in a predominantly primer-independent manner (33,35) (Fig. 4C). Consistent with the native gel analyses, the ratio between short and long RNA products was noticeably higher for
QDE-1\(^{Te}\) than for QDE-1\(^{Ncr}\) (Fig. 4C). Importantly, the bias for short RNA products was consistently detected for several independently purified batches of QDE-1\(^{Te}\) and observed over a wide temperature range (Fig. 4A-B). This ruled out the trivial explanation that the enzymatic differences between QDE-1\(^{Te}\) and QDE-1\(^{Ncr}\) were related to a higher growth temperature limit of \(T.\) terrestris compared to \(N.\) crassa (60-62). In fact, the thermal stability of the QDE-1\(^{Te}\) protein exceeded that of QDE-1\(^{Ncr}\) by only 3.7°C (Tm=54.0°C±0.0 vs. Tm=50.3°C±0.6; Fig. 5B). The two polymerases also had comparable single-stranded template binding properties (Figs 5C and D). Thus, distinct QDE-1 orthologs may have markedly different RNA polymerization properties.

Evolutionary divergence between QDE-1 family members modulates their product preferences—To ensure that functional differences between QDE-1\(^{Ncr}\) and QDE-1\(^{Te}\) resulted from evolutionary innovation, we purified a catalytic fragment of QDE-1 from \(Myceliophthora\) thermophila, a Chaetomiaceae fungus more closely related to \(T.\) terrestris than to \(N.\) crassa (Figs. 1 and 5). The Tm of QDE-1\(^{Mth}\) was comparable to that of QDE-1\(^{Ncr}\) and QDE-1\(^{Te}\) (Tm=52.3°C±0.6 vs. Tm=50.3°C±0.6 and Tm=54.0°C±0.0, respectively; Fig. 5).

To compare enzymatic properties of QDE-1\(^{Ncr}\), QDE-1\(^{Te}\) and QDE-1\(^{Mth}\), we incubated the three polymerases with either ssRNA or ssDNA template for 1 hour at 45°C and separated the reaction products by urea-polyacrylamide gel electrophoresis (PAGE) affording simultaneous detection of long RNAs migrating at the top of the lane and sRNA products visualized at a single-nucleotide resolution (Fig. 6). In ssRNA-programmed reactions, QDE-1\(^{Mth}\) polymerase was clearly more efficient in producing short RNA products than QDE-1\(^{Ncr}\), albeit not to the same extent as QDE-1\(^{Te}\) (Fig. 6A). Moreover, QDE-1\(^{Ncr}\) synthesized detectable amounts of ≤23-nt and ~28-31-nt products but virtually no 24-27-nt long sRNAs (Fig. 6A). On the other hand, both QDE-1\(^{Te}\) and QDE-1\(^{Mth}\) efficiently produced 24-27-nt sRNA in addition to shorter products (Fig. 6A). Interestingly, each of the three polymerases generated a unique blend of sRNA products in ssDNA-programmed reactions (Fig. 6B). As an additional control, we purified a D607A mutant QDE-1\(^{Mth}\), where the first Asp residue of the DYDGD motif was mutated to Ala. As expected, the mutant protein lacked detectable nucleotidyl transferase activity (Fig. 7).

We concluded that sequence divergence between QDE-1 orthologs appears to underlie differences in their activities.

QDE-1\(^{Te}\) is structurally similar to QDE-1\(^{Ncr}\)—We next wondered whether distinct functional properties of QDE-1 orthologs might be due to major differences in spatial structures of these enzymes. This appeared plausible given the degree of amino acid sequence divergence outside of the active center (see Fig. 2). To this end, we determined QDE-1\(^{Te}\) 3D structure using X-ray crystallography. A 3.19 Å-resolution trace of the QDE-1\(^{Te}\) polypeptide chain revealed two closely similar QDE-1\(^{Ncr}\)-like homodimers in the asymmetric unit (r.m.s. deviation of 1.06 Å between dimer A/B and C/D for 917 α-carbon atoms). In each dimer, the two monomers are related by a non-crystallographic dyad (Fig. 8A).

Other structural details QDE-1\(^{Te}\) also resembled those of QDE-1\(^{Ncr}\) (40). For example, the QDE-1\(^{Te}\) dimer had a pyramidal shape with a pair of double-psi beta-barrel domains (DPBB1 and DPBB2; 312-413 and 535-639) located at the base of each subunit (Fig. 8A-B). The “head” (residues 457-507 and 822-1006 of the catalytic fragment) and the “neck” domains (residues 428-532 and 784-821) of the two subunits adopted a slightly tilted back-to-back position (Fig. 8A) and the “slab” domain (residues 1-244) of each subunit protruded from the catalytic
domain towards the corresponding head domain. The five domains formed an extensive groove likely accommodating the template and the nascent RNA product, with the catalytic site located at the bottom (Fig. 8A-B). In general, Cα atoms in the QDE-1\(^{Ncr}\) and QDE-1\(^{Te}\) dimers superimposed very well with a r.m.s.d score of 1.34 Å for 917 α-carbon atoms (Fig. 8C).

As previously found for QDE-1\(^{Ncr}\), one of the two QDE-1\(^{Te}\) subunits adopted a slightly more closed conformation than the other. Superposition of 922 Cα atoms between the two monomers (dimer A/B) using the SSM server (www.ebi.ac.uk/msd-srv/ssm) returned a r.m.s.d. of 2.15 Å (Fig. 8D). Largest conformational differences between the two monomers occurred closer to the C-terminal end of the polypeptide chain. The homodimer interface is stabilized by a total of 119 and 115 residues of monomer A and B, respectively (63). Most residues at the interface originate from the upper “neck” and “head” domains, where the subunits display the largest structural differences (Fig. 8C). These data suggest that QDE-1\(^{Ncr}\) and QDE-1\(^{Te}\) have remarkably similar tertiary and quaternary structures despite their primary sequence differences.

\(\text{QDE-1}^{\text{Te}}\) can function as a dimer in solution—Extensive contacts between the A and B subunits observed in the QDE-1\(^{Ncr}\) and QDE-1\(^{Te}\) crystals suggested that these proteins may form catalytically active dimers in solution with the two monomers oscillating between the closed and open conformations. To test this possibility directly, we used electron microscopy (EM) to compare negatively stained images of the QDE-1\(^{Te}\) apoenzyme and its complex with a ssDNA template (Fig. 9), using reference-free 2D class averaging (Fig. 9B). These initial EM reconstructions carried out using 3,000 apoenzyme and 3,000 ssDNA bound particles revealed largely similar pyramid-shaped homodimer structures each containing two quasi-symmetric grooves and showing at this resolution a generally good fit with the QDE-1\(^{Te}\) crystal structure obtained without bound DNA (Fig. 9C). Like in the crystal structure, one of the two template/product grooves in the QDE-1\(^{Te}\) dimer adopted a more open conformation in the EM reconstructions (Fig. 9C). To obtain a better-resolved solution structure of QDE-1\(^{Te}\) in the presence of ssDNA, we extended our EM data analysis to include a total of 11,992 particles. This allowed us to visualize this complex to a resolution of 20 Å (Fig.10 A-B). As expected, the structural fit between the EM and crystal structures of QDE-1\(^{Te}\) further improved when the EM map was prepared using the crystal structure as a reference (Fig. 10B). Of note, even this improved resolution was not sufficiently high to unambiguously assign density for the ssDNA. Likewise, understanding of the subtle conformational changes in the QDE-1\(^{Te}\) dimer upon template binding will require a higher resolution structure of the RdRP-ssDNA complex, a challenge that will be addressed in the future. Nonetheless, our present data strongly suggest that QDE-1 orthologs function as dimers in solution.

DISCUSSION

This study suggests that fungal QDE-1 orthologs underwent an unusually rapid sequence divergence outside of the catalytic center and the nucleotide-binding site. This corresponds to a significantly higher amino acid substitution rate in this group compared to its RdRP paralogs and distant DdRP relatives (Fig. 2). Consistent with their stronger divergence from QDE-1\(^{Ncr}\) than from each other (Fig. 1), QDE-1\(^{Te}\) and QDE-1\(^{Mth}\) differ from QDE-1\(^{Ncr}\) in their enhanced ability to generate sRNA copies and reduced ability to synthesize long RNA products (Fig. 4). Since QDE-1\(^{Ncr}\), QDE-1\(^{Te}\) and QDE-1\(^{Mth}\) originate from a single taxonomic order (Sordariales), these data argue that protein
functions can undergo substantial changes over relatively short periods of evolutionary history. Notably, this effect was not due to a recent gene duplication event followed by functional specialization of the newly emerged paralogs since our BLAST searches did not reveal any additional RdRP gene in the *N. crassa*, *T. terrestris* and *M. thermophila* genomes besides QDE-1, SAD-1 and RRP-3.

What could be biological significance of such accelerated functional evolution? One possible answer relates to the role of RNA silencing in cellular defense against viruses and transposable elements (3,5,6). Different species encounter distinct sets of pathogens, which necessitate corresponding changes in cellular defense mechanisms (64). Interestingly, *N. crassa* is a mesophilic fungus colonizing burnt vegetation and occasionally living plants (60,61), whereas *T. terrestris* and *M. thermophila* typically inhabit self-heating composts (62). Exposure to diverse environments and correspondingly different sets of biohazards might have exerted substantial evolutionary pressure on cellular immunity factors including QDE-1. An interesting direction for future studies would be to test if this also rewired RNA silencing pathways in a more fundamental manner, e.g. incorporating sRNA products of QDE-1<sup>Tte</sup> and QDE-1<sup>Mth</sup> into signal amplification loops similar to the secondary siRNA pathway in *C. elegans* (16,19-21,39).

The striking structural similarity between QDE-1<sup>Tte</sup> and QDE-1<sup>Ncr</sup> uncovered in this work (Fig. 8) suggests that evolution in this group likely proceeded through alteration of functionally important surface residues rather than major changes in the protein fold. In other words, QDE-1 orthologs show a combination of structural robustness and functional innovability (65). Although further work will be required to identify specific structural elements underlying functional divergence in the QDE-1 group, these will likely include amino acid residues directly interacting with the template-product duplex or affecting protein flexibility. Indeed, structural alignment of QDE-1<sup>Ncr</sup> or QDE-1<sup>Tte</sup> with the yeast pol-II elongation complex [PDB 1R9T; (51)] using structural similarity between QDE-1 DPBB2 domain and the only DPBB domain of the pol-II RPB1 subunit suggests that QDE-1 enzyme may have to undergo major conformational changes to allow egress of template-product duplexes longer than ~10 base pairs.

The EM reconstructions presented here provide an unprecedented insight into solution structure of an RdRP enzyme (Figs. 9 and 10). In line with the earlier prediction (40), QDE-1 assembles into a pyramid-shape homodimer with each of the two subunits containing a groove suitable for template/product binding. One of the two subunits in the EM density maps adopts a more “closed” conformation than the other. This supports the “two-stroke motor” model of QDE-1 activity that was proposed earlier based on the X-ray structure of QDE-1<sup>Ncr</sup> apoenzyme (40). Importantly, our data indicate that QDE-1 retains its dimeric form in the presence of a single-stranded DNA template (Figs. 9 and 10).

We finally note that the efficient synthesis of sRNA products by the newly isolated QDE-1 enzymes might facilitate a range of research and diagnostic applications. QDE-1<sup>Tte</sup> appears to generate more than one copy of a given template sequence (Fig. 4A-B). Therefore, this enzyme might be especially useful for amplifying either an entire nucleic acid target or its parts accessible to the polymerase. Combined with deep sequencing technology, this may open up new possibilities in high-throughput analyses of the transcriptome composition, RNA conformation and ribonucleoprotein complex structure. Enzymatic properties of QDE-1<sup>Tte</sup> could be further improved by knowledge-based mutagenesis of its evolutionarily variable parts or by “shuffling” corresponding
sequences with their counterparts from other QDE-1 orthologs.

In conclusion, our study argues that acquisition of novel enzymatic properties through divergence of orthologous sequences could be a more common evolutionary scenario than anticipated previously. This work also improves our understanding of molecular mechanisms underlying RdRP functions and expands the existing molecular biology toolkit. We predict that further comparative analyses of this remarkably diverse class of enzymes will be a rewarding experience for evolutionary biologists and biochemists alike.

Acknowledgments: We thank Andrew Wong for technical help with EM and Bernard Henrissat and Snezhka Oliferenko for helpful discussions.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: XQ and DAD expressed the proteins (QDE-1\textsuperscript{Ncr}, QDE-1\textsuperscript{Mth}, QDE-1\textsuperscript{Te} and QDE-1\textsuperscript{Mth} (D607A)). XQ purified the proteins, conducted the thermofluor assay, crystallized QDE-1\textsuperscript{Te} and solved its structure. XQ and FMH conducted the polymerase activity assay. WYH generated the mutated polymerase QDE-1\textsuperscript{Mth} (D607A). SB conducted the EM and refined the EM structure of QDE-1\textsuperscript{Te}. FMH and EVM conducted phylogenetic analysis. JL collected the diffraction data and, with AES, refined the crystal structure of QDE-1\textsuperscript{Te}. XQ, EVM and JL wrote the paper.

REFERENCES


FOOTNOTES
This work was supported by the Ministry of Education Singapore (Academic Research Fund Tier 1 grant RG62/12; J.L.), National Medical Research Council (NMRC/ CBRG/0028/2013; E.V.M.), Biotechnology and Biological Sciences Research Council (BB/M001199/1 and BB/M007103/1; E.V.M.) and the Nanyang President Graduate Scholarship (F.M.H.).

The abbreviations used are: RdRP, RNA dependent RNA polymerases; ssRNA, small RNA; ds, double-stranded; ss, single-stranded; aRNA, aberrant ssRNA; ORF, open reading frame; TCEP, tris(2-carboxyethyl)phosphine; TCA, trichloroacetic acid, EM, electron microscopy, TEM, transmission electron microscopy; RMSD, root-mean-square deviation; FSC, Fourier shell correlation.

FIGURE LEGENDS

Figure 1. Phylogenetic analysis of fungal QDE-1, SAD1 and RRP-3 RdRPs. QDE-1 and SAD-1 orthologs were detected in all 37 Ascomycota species used in this analysis and RRP-3 was present in 34 species. The tree was constructed using the Neighbor-Joining method and drawn to scale with the evolutionary distances computed as numbers of amino acid substitutions per site. All positions with less than 50% site coverage were eliminated. Percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Neurospora crassa RdRPs are highlighted in red. The tree additionally shows Schizosaccharomyces pombe and Schizosaccharomyces japonicus fission yeast Rdr1 sequences (UniProt IDs RDR1_SCHPO and B6K409_SCHJY, respectively) clustering with the SAD-1 RdRP branch and the RPB1 subunit of the Saccharomyces cerevisiae pol-II DdRP used as an outgroup (UniProt ID RPB1_YEAST).

Figure 2. QDE-1 orthologs rapidly diverge in evolution. A. Similarity plots for the QDE-1, SAD-1 and RRP-3 orthologs as well as for RPB1 and RPB2 (bottom) subunits of pol-II DdRP from 37 fungal species containing more than one paralogous RdRP. Red arrowheads mark positions of the catalytic D(Y/F/L)DGD motif in the RdRP and RPB1 active centers (RPB2 lacks this sequence). The gray dimension line indicates the C-terminal part of QDE-1 previously shown to be enzymatically active (33). Note that the DYDGD-adjacent region in the QDE-1 orthologs is conserved noticeably better than the rest of the sequence. This contrasts with the rest
of the proteins containing wider areas of relatively strong conservation. B. Box and whisker plot comparison of amino acid substitution scores showing significantly higher divergence rate in QDE-1 proteins compared to their SAD-1, RRP-3, RPB1 and RPB2 counterparts. Corresponding Wilcoxon test p-values are indicated on the top. C. Surfaces proximal (≤4 Å) to the template (dark gray) and the nascent RNA product (black) in the “closed” subunit of Neurospora crassa QDE-1 (QDE-1\textsuperscript{Ncr}) [PDB 2J7N; (40)] and S. cerevisiae pol-II [PDB 1R9T; (51)]. Incoming ATP monomer and the two catalytic ions, Mg\textsuperscript{2+} and Mg\textsuperscript{2+}, are shown in black and green, respectively. Protein surfaces are colored using cyan for low, white for intermediate and maroon for high conservation. In the case of QDE-1\textsuperscript{Ncr}, Mg\textsuperscript{2+} position is determined experimentally, whereas template, RNA product, incoming ATP and Mg\textsuperscript{2+} are modeled in based on the pol-II elongation structure on the right.

Figure 3. Spatial clustering of evolutionarily conserved amino acid sequences in QDE-1\textsuperscript{Ncr} structure. Surface representations of the QDE-1\textsuperscript{Ncr} protein homodimer show the “closed” A subunit colored according to interspecies conservation and the “open” B subunit colored in beige. Known position of Mg\textsuperscript{2+} and predicted positions of the second catalytic Mg\textsuperscript{2+} ion (Mg\textsuperscript{2+}), an incoming ATP, a template and an RNA product are also indicated. Note that highly conserved amino acid residues cluster in the vicinity of the catalytic center and the NTP pore.

Figure 4. QDE-1 orthologs have distinct enzymatic properties. A-B. RNA-polymerase activities of purified catalytic fragments of QDE-1\textsuperscript{Ncr} and QDE-1\textsuperscript{Tte} were assayed at 30-60°C in the presence of (A) a ssRNA or (B) a ssDNA template. The reaction products were separated by native agarose gel electrophoresis and visualized using ethidium bromide staining or 32P phosphorimaging, as indicated. Note that QDE-1\textsuperscript{Tte} is substantially more efficient than QDE-1\textsuperscript{Ncr} in generating sRNA products that migrate either at a low-molecular-weight position or in a template base-paired form. C. RNA-polymerase activities of QDE-1\textsuperscript{Ncr} and QDE-1\textsuperscript{Tte} were assayed at 45°C in the presence of a ssRNA template and analyzed by denaturing agarose gel electrophoresis. D. RNA products from (C) were incubated with increasing concentrations of RNase ONE or RNase ONE reaction buffer, as specified under Experimental Procedures. Positions of the 1× full template-length products of processive end-to-end polymerization initiated in a primer-independent manner and sRNA products of non-processive polymerization are indicated on the right. Also shown is an expected position of “back-primed” 2× template-length products, which QDE-1\textsuperscript{Ncr} can generate for some but not all ssRNA templates (33).

Figure 5. Purification and biochemical characterization of QDE-1 orthologs. A. SDS-PAGE analysis of purified recombinant QDE-1\textsuperscript{Ncr}, QDE-1\textsuperscript{Tte} and QDE-1\textsuperscript{Mth} catalytic fragments. B. Thermostability analyses showing that the three proteins have comparable melting temperatures, 50.3, 54.0 and 52.3°C, respectively. C-D. Binding of a short ssDNA template by QDE-1\textsuperscript{Ncr} and QDE-1\textsuperscript{Tte} was analyzed using a slot blot filter-binding assay as described in Experimental Procedures. Binding data in C are averaged from three independent experiments ±SE and the inset shows an example of filter binding data for 1 µM ssDNA incubated with QDE-1\textsuperscript{Ncr}, QDE-1\textsuperscript{Tte} or BSA control. Panel D shows a Lineweaver-Burk plot of the data in (C) used to determine apparent K\textsubscript{d} values for QDE-1\textsuperscript{Ncr} and QDE-1\textsuperscript{Tte}.

Figure 6. QDE-1 orthologs synthesize different blends of sRNA products. A-B. Purified QDE-1\textsuperscript{Ncr} and QDE-1\textsuperscript{Tte} and QDE-1\textsuperscript{Mth} were assayed at 45°C with (A) a ssRNA or (B) a ssDNA
template and analyzed by urea-containing 15% polyacrylamide gel electrophoresis. Note that in
the presence of a long ssRNA, QDE-1Tte and QDE-1Mth generate a relatively higher fraction of
sRNAs than QDE-1Ncr, especially those of 12 nt and 24-27 nt. On the other hand, QDE-1Ncr
synthesizes a relatively larger amount of long RNA products migrating close to the top of the
lane. Each of the 3 polymerases produces a unique combination of sRNA products from the
ssDNA. In the lane scans provided at the bottom of each panel, maximal intensity of the 18-nt
marker band is set to 1.

Figure 7. Lack of RNA polymerase activity in the QDE-1Mth (D607A) mutant.
The QDE-1Mth (D607A) mutant containing the AYDGD sequence instead of the wild-type
catalytic DYDGD motif shows no RNA polymerase activity in the presence of either ssRNA or
ssDNA template and over a wide range of reaction temperatures.

Figure 8. QDE-1Tte crystal structure. A. QDE-1Tte is an asymmetric dimer with each subunit
containing a DPBB1 (shown in cyan for subunit A), a DPBB2 (green), a “head” (magenta), a
“neck” (blue), and a “slab” domain (red). The entire B subunit is colored in gray B. Magnified
view of the DPBB1 and DPBB2 domains with the three catalytic aspartate side chains shown as
balls and sticks. C. Structural alignment between QDE-1Tte and previously solved QDE-1Ncr
structure showing a considerable overlap between the two proteins. D. Structural alignment
between subunits A and B QDE-1Tte demonstrating that subunit A adopts a more “closed”
conformation than subunit B. The two subunits are color-coded as in panel (A).

Figure 9. Electron microscopy reconstructions of the apo- and ssDNA-bound QDE-1Tte
structures. A. A total of 3,000 particles were used for the analysis presented in this figure. A
negatively stained electron micrograph showing particles of QDE-1Tte dimers in complex with a
ssDNA template. B. Representative reference-free 2D class averages of the particles in (A). C.
Overlay of QDE-1Tte structures determined by reference-free EM reconstruction in the presence
(or absence of a ssDNA template (gray). The crystal structure of the unbound QDE-1Tte
dimer is overlaid, with its subunits colored in cyan and gray respectively. Note that at this
resolution of about 31 Å (1/0.032 Å³ shown by the Fourier shell correlation map as inset, both
the apoenzyme and the ssDNA complex appear as structurally similar dimers.

Figure 10. Higher resolution EM structures of the QDE-1Tte ssDNA complex. EM data
analysis using a total of 11,992 particles is shown for the reference-free QDE-1Tte (panel A) or
using the crystal structure of unbound QDE-1Tte as a reference (panel B). The crystal structure of
QDE-1Tte was fitted into each of the EM reconstructions. For each reconstruction, a front and
back view is provided with subunits of QDE-1Tte rendered in cyan and gray. Panel (A)
additionally shows the Fourier Shell Correlation (FSC) curve from the final refined density map
of QDE-1Tte-ssDNA. In both reconstructions the resolution is ~20 Å (1/0.051 Å⁻¹).
Table 1. Primers used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence, 5’ to 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>QDE1_Ncr_forward</td>
<td>TACTTCCAATCCATGGCTGGCTCGGAGCGAAGAAA</td>
</tr>
<tr>
<td>QDE1_Ncr_reverse</td>
<td>TATCCACCTTTACTGTAATACGTCCATTCCTGGAA</td>
</tr>
<tr>
<td>QDE1_Mth_forward</td>
<td>TACTTCCAATCCATGGATCACTTCAGACTGC</td>
</tr>
<tr>
<td>QDE1_Mth_reverse</td>
<td>TATCCACCTTTACTGTCATTACTCGTCACTCGGAAATCGC</td>
</tr>
<tr>
<td>QDE1_Tte_forward</td>
<td>TACTTCCAATCCATGGGATGACCGCCGCCTC</td>
</tr>
<tr>
<td>QDE1_Tte_reverse</td>
<td>TATCCACCTTTACTGTCACCGGTCACCCGCG</td>
</tr>
<tr>
<td>D607A_forward</td>
<td>GACAAGTTGTCAGGTGGTGCCTACGATGGAG</td>
</tr>
<tr>
<td>D607A_reverse</td>
<td>GGCGAAGTCTCCATCGTAGGACCACCTGAC</td>
</tr>
</tbody>
</table>
## Table 2. Data collection and refinement statistics for QDE-1<sup>Tth</sup> apoenzyme.

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Detector Type</td>
<td>Pilatus 6M</td>
</tr>
<tr>
<td>Synchrotron</td>
<td>SLS PXIII</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.03320</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>48.91 – 3.19 (3.38 – 3.19)</td>
</tr>
<tr>
<td>Space Group</td>
<td>P2&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>Dimers per asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td>Cell parameters (Å)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>84.23, 165.84, 173.83 90, 90.10, 90</td>
</tr>
<tr>
<td>Measured reflections&lt;sup&gt;a&lt;/sup&gt;</td>
<td>243,519 (38,635)</td>
</tr>
<tr>
<td>Unique reflections&lt;sup&gt;a&lt;/sup&gt;</td>
<td>77,595 (12,236)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.35 (1.6)</td>
</tr>
<tr>
<td>I/σ&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.8 (3.0)</td>
</tr>
<tr>
<td>Completeness (%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>98.6 (95.5)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.169 (0.399)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement statics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflections used for refinement</td>
<td>75,503</td>
</tr>
<tr>
<td>Non-hydrogen atoms&lt;sup&gt;b&lt;/sup&gt;</td>
<td>28,130</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; (%)</td>
<td>21.30 (25.06)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>25.09 (26.20)</td>
</tr>
<tr>
<td>r.m.s.d. bond lengths (Å)</td>
<td>0.015</td>
</tr>
<tr>
<td>r.m.s.d. bond angles (°)</td>
<td>1.85</td>
</tr>
<tr>
<td>Mean B-factor (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>60.58 (16.32, 33.99, 64.8, 77.5)</td>
</tr>
<tr>
<td>Ramachandran Plot</td>
<td></td>
</tr>
<tr>
<td>Most favored (%)</td>
<td>83.2</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>13.8</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>1.4</td>
</tr>
</tbody>
</table>

PDB code 5FSW

<sup>a</sup>The values for the highest resolution shell are shown in parenthesis.

<sup>b</sup>Residues missing from the model are: 213-225, 566-567, 813-816, 867-875, 895-906, 923-925, 967-972 and 1007-1034.

<sup>c</sup>The B-factor values are given for all the protein atoms and for each of the four polypeptide chains in the asymmetric unit.
<table>
<thead>
<tr>
<th>Order</th>
<th>Genus</th>
<th>species</th>
<th>QDE-1</th>
<th>SAD-1</th>
<th>RRP-3</th>
<th>RP1B</th>
<th>RP2B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botryosphaeriales</td>
<td>Macrophomina</td>
<td>phaseolina</td>
<td>K2RD90_MACPH</td>
<td>K2QN15_MACPH</td>
<td>K2RKL1_MACPH</td>
<td>K2R7EZ_MACPH</td>
<td>K2RUI9_MACPH</td>
</tr>
<tr>
<td>Capnidioidales</td>
<td>Bauldella</td>
<td>companiensis</td>
<td>M2N190_BAUO</td>
<td>M2N260_BAUO</td>
<td>M2N323_BAUO</td>
<td>M2N508_BAUO</td>
<td>M2N114_BAUO</td>
</tr>
<tr>
<td>Capnidioidales</td>
<td>Pseudocorcorina</td>
<td>sphaerella</td>
<td>M3B66_Y6_SPHMS</td>
<td>M3CB8_L8_SPHMS</td>
<td>M3AB5_SPHMS</td>
<td>M3AC5_SPHMS</td>
<td>M3AC5_SPHMS</td>
</tr>
<tr>
<td>Capnidioidales</td>
<td>Sphaerulina</td>
<td>musivia</td>
<td>M3CWL8_SPHMS</td>
<td>M3CB5_L8_SPHMS</td>
<td>M3AB5_SPHMS</td>
<td>M3AC5_SPHMS</td>
<td>M3AC5_SPHMS</td>
</tr>
<tr>
<td>Capnidioidales</td>
<td>Zymosporia</td>
<td>tritici</td>
<td>F9X324_MYCMG</td>
<td>F9X324_MYCMG</td>
<td>F9X224_MYCMG</td>
<td>F9X224_MYCMG</td>
<td>F9X224_MYCMG</td>
</tr>
<tr>
<td>Chaetothyriales</td>
<td>Conidiopora</td>
<td>apolinis</td>
<td>R74Y14_CONA1</td>
<td>R74Y14_CONA1</td>
<td>R74Y14_CONA1</td>
<td>R74Y14_CONA1</td>
<td>R74Y14_CONA1</td>
</tr>
<tr>
<td>Eurotiales</td>
<td>Aspergillus</td>
<td>niger</td>
<td>G3Y305_20ASNA</td>
<td>G3Y305_20ASNA</td>
<td>G3Y305_20ASNA</td>
<td>G3Y305_20ASNA</td>
<td>G3Y305_20ASNA</td>
</tr>
<tr>
<td>Eurotiales</td>
<td>Aspergillus</td>
<td>oryzae</td>
<td>I8TQ98_AP20</td>
<td>I8TQ98_AP20</td>
<td>I8TQ98_AP20</td>
<td>I8TQ98_AP20</td>
<td>I8TQ98_AP20</td>
</tr>
<tr>
<td>Eurotiales</td>
<td>Penicillium</td>
<td>chryogenous</td>
<td>B6HH83_PENCW</td>
<td>B6HH83_PENCW</td>
<td>B6HH83_PENCW</td>
<td>B6HH83_PENCW</td>
<td>B6HH83_PENCW</td>
</tr>
<tr>
<td>Eurotiales</td>
<td>Penicillium</td>
<td>digitatum</td>
<td>K6FW5_M_PENDI</td>
<td>K6FW5_M_PENDI</td>
<td>K6FW5_M_PENDI</td>
<td>K6FW5_M_PENDI</td>
<td>K6FW5_M_PENDI</td>
</tr>
<tr>
<td>Glomerellales</td>
<td>Colletotrichum</td>
<td>glaucosporeides</td>
<td>L2FZ96_COLON</td>
<td>L2FZ96_COLON</td>
<td>L2FZ96_COLON</td>
<td>L2FZ96_COLON</td>
<td>L2FZ96_COLON</td>
</tr>
<tr>
<td>Glomerellales</td>
<td>Colletotrichum</td>
<td>graninocilum</td>
<td>E3Q45_COLON</td>
<td>E3Q45_COLON</td>
<td>E3Q45_COLON</td>
<td>E3Q45_COLON</td>
<td>E3Q45_COLON</td>
</tr>
<tr>
<td>Helotiales</td>
<td>Botryosina</td>
<td>fucellina</td>
<td>M7U3R5_BO2F1</td>
<td>M7U3R5_BO2F1</td>
<td>M7U3R5_BO2F1</td>
<td>M7U3R5_BO2F1</td>
<td>M7U3R5_BO2F1</td>
</tr>
<tr>
<td>Helotiales</td>
<td>Sclerotinia</td>
<td>sclerotiorum</td>
<td>A7ED71_SCLS1</td>
<td>A7ED71_SCLS1</td>
<td>A7ED71_SCLS1</td>
<td>A7ED71_SCLS1</td>
<td>A7ED71_SCLS1</td>
</tr>
<tr>
<td>Hypocreales</td>
<td>Claviceps</td>
<td>purpurea</td>
<td>M1WA51_CLAP2</td>
<td>M1WA51_CLAP2</td>
<td>M1WA51_CLAP2</td>
<td>M1WA51_CLAP2</td>
<td>M1WA51_CLAP2</td>
</tr>
<tr>
<td>Hypocreales</td>
<td>Fusarium</td>
<td>gramineum</td>
<td>H1R004_GIBZE</td>
<td>H1R004_GIBZE</td>
<td>H1R004_GIBZE</td>
<td>H1R004_GIBZE</td>
<td>H1R004_GIBZE</td>
</tr>
<tr>
<td>Hypocreales</td>
<td>Fusarium</td>
<td>oxyssporum</td>
<td>F9XJ2_FUSOF</td>
<td>F9XJ2_FUSOF</td>
<td>F9XJ2_FUSOF</td>
<td>F9XJ2_FUSOF</td>
<td>F9XJ2_FUSOF</td>
</tr>
<tr>
<td>Hypocreales</td>
<td>Nectria</td>
<td>haematococca</td>
<td>C7Z29_NEC17</td>
<td>C7Z29_NEC17</td>
<td>C7Z29_NEC17</td>
<td>C7Z29_NEC17</td>
<td>C7Z29_NEC17</td>
</tr>
<tr>
<td>Hypocreales</td>
<td>Trichoderma</td>
<td>atroviride</td>
<td>G9P762_HYPAI</td>
<td>G9P762_HYPAI</td>
<td>G9P762_HYPAI</td>
<td>G9P762_HYPAI</td>
<td>G9P762_HYPAI</td>
</tr>
<tr>
<td>Hypocreales</td>
<td>Trichoderma</td>
<td>vires</td>
<td>G9N8C5_HYHYV</td>
<td>G9N8C5_HYHYV</td>
<td>G9N8C5_HYHYV</td>
<td>G9N8C5_HYHYV</td>
<td>G9N8C5_HYHYV</td>
</tr>
<tr>
<td>Magnaporthiales</td>
<td>Magnaporia</td>
<td>poae</td>
<td>M4F989_MAGP6</td>
<td>M4F989_MAGP6</td>
<td>M4F989_MAGP6</td>
<td>M4F989_MAGP6</td>
<td>M4F989_MAGP6</td>
</tr>
<tr>
<td>Orygenales</td>
<td>Cocidiodes</td>
<td>inimis</td>
<td>J2K906_COCIM</td>
<td>J1K906_COCIM</td>
<td>J1K906_COCIM</td>
<td>J1K906_COCIM</td>
<td>J1K906_COCIM</td>
</tr>
<tr>
<td>Orygenales</td>
<td>Cocidiodes</td>
<td>nodosus</td>
<td>E9DB6_L6_COPPS</td>
<td>E9DB6_L6_COPPS</td>
<td>E9DB6_L6_COPPS</td>
<td>E9DB6_L6_COPPS</td>
<td>E9DB6_L6_COPPS</td>
</tr>
<tr>
<td>Orygenales</td>
<td>Paracocidioides</td>
<td>brasiliensis</td>
<td>C1G6K3_PARDB</td>
<td>C1G6K3_PARDB</td>
<td>C1G6K3_PARDB</td>
<td>C1G6K3_PARDB</td>
<td>C1G6K3_PARDB</td>
</tr>
<tr>
<td>Ophiostomatiales</td>
<td>Ophiostoma</td>
<td>piceae</td>
<td>S3CA4_OPIH1</td>
<td>S3CA4_OPIH1</td>
<td>S3CA4_OPIH1</td>
<td>S3CA4_OPIH1</td>
<td>S3CA4_OPIH1</td>
</tr>
<tr>
<td>Ophiostomatiales</td>
<td>Ophiostoma</td>
<td>schenckii</td>
<td>U7P651_SPOS1</td>
<td>U7P651_SPOS1</td>
<td>U7P651_SPOS1</td>
<td>U7P651_SPOS1</td>
<td>U7P651_SPOS1</td>
</tr>
<tr>
<td>Pleosporales</td>
<td>Phaeospora</td>
<td>nodosum</td>
<td>Q6V296_PHANO</td>
<td>Q6V296_PHANO</td>
<td>Q6V296_PHANO</td>
<td>Q6V296_PHANO</td>
<td>Q6V296_PHANO</td>
</tr>
<tr>
<td>Pleosporales</td>
<td>Setosphaeria</td>
<td>turgida</td>
<td>R04K68_SET2T</td>
<td>R04K68_SET2T</td>
<td>R04K68_SET2T</td>
<td>R04K68_SET2T</td>
<td>R04K68_SET2T</td>
</tr>
<tr>
<td>Saccharomycetaceae</td>
<td>Saccharomyces</td>
<td>cerevisiae</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>RPB1_20YEAST</td>
</tr>
<tr>
<td>Schizophilomyces</td>
<td>Schizophilomyces</td>
<td>pombe</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>RPB2_20YEAST</td>
</tr>
<tr>
<td>Schizophilomyces</td>
<td>Schizophilomyces</td>
<td>japonicus</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>RPB2_20YEAST</td>
</tr>
<tr>
<td>Sordariales</td>
<td>Chaetomum</td>
<td>thermophilum</td>
<td>G0S882_Y7_CHATD</td>
<td>G0S882_Y7_CHATD</td>
<td>G0S882_Y7_CHATD</td>
<td>G0S882_Y7_CHATD</td>
<td>G0S882_Y7_CHATD</td>
</tr>
<tr>
<td>Sordariales</td>
<td>Myceliophilum</td>
<td>thermophilum</td>
<td>G2Q4203_20THA</td>
<td>G2Q4203_20THA</td>
<td>G2Q4203_20THA</td>
<td>G2Q4203_20THA</td>
<td>G2Q4203_20THA</td>
</tr>
<tr>
<td>Sordariales</td>
<td>Neurospora</td>
<td>crassa</td>
<td>K19D64_NEURC</td>
<td>K19D64_NEURC</td>
<td>K19D64_NEURC</td>
<td>K19D64_NEURC</td>
<td>K19D64_NEURC</td>
</tr>
<tr>
<td>Sordariales</td>
<td>Neurospora</td>
<td>tetrasperma</td>
<td>F8N9K5_NEUT8</td>
<td>F8N9K5_NEUT8</td>
<td>F8N9K5_NEUT8</td>
<td>F8N9K5_NEUT8</td>
<td>F8N9K5_NEUT8</td>
</tr>
<tr>
<td>Sordariales</td>
<td>Sordaria</td>
<td>macrospora</td>
<td>F7SVYI2_SORMK</td>
<td>F7SVYI2_SORMK</td>
<td>F7SVYI2_SORMK</td>
<td>F7SVYI2_SORMK</td>
<td>F7SVYI2_SORMK</td>
</tr>
<tr>
<td>Sordariales</td>
<td>Thalasia</td>
<td>terrresists</td>
<td>G2R9111_HFITE</td>
<td>G2R9111_HFITE</td>
<td>G2R9111_HFITE</td>
<td>G2R9111_HFITE</td>
<td>G2R9111_HFITE</td>
</tr>
</tbody>
</table>
Functional evolution of fungal RNA polymerases

Fig. 1
Functional evolution of fungal RNA polymerases

Fig. 2
Functional evolution of fungal RNA polymerases

Fig. 3
Fig. 4
Functional evolution of fungal RNA polymerases

Fig. 5
Functional evolution of fungal RNA polymerases

Fig. 6
Fig. 7
Functional evolution of fungal RNA polymerases

Fig. 8

A. QDE-1<sup>Tfo</sup> “closed” subunit A
B. QDE-1<sup>Tfo</sup> “open” subunit B
C. Structural alignment between QDE-1<sup>Tfo</sup> and QDE-1<sup>Ncr</sup>
D. Structural alignment between QDE-1<sup>Tfo</sup> subunit A and B
Functional evolution of fungal RNA polymerases

Fig. 9
Fig. 10